Methods:

Clinical Recruitment and Sample Collection

61 individuals (45 SARS-CoV-2 naïve, 16 SARS-CoV-2 recovered) were consented and enrolled in the longitudinal vaccine study with approval from the University of Pennsylvania Institutional Review Board (IRB# 844642). All participants were otherwise healthy and based on self-reported health screening did not have any history of chronic health conditions. Subjects were stratified based on self-reported and laboratory evidence of a prior SARS-CoV-2 infection. All subjects received either Pfizer (BNT162b2) or Moderna (mRNA-1273) mRNA vaccines. Samples were collected at 6 timepoints: baseline, ~2 weeks post-primary immunization, day of secondary immunization, ~1 week post-secondary immunization, ~3 months post-primary immunization, and ~6 months post-primary immunization. 80-100mL of peripheral blood samples and clinical questionnaire data were collected at each study visit. A separate cohort of 26 SARS-CoV-2 convalescent individuals was used to compare vaccine-induced immune responses to immune responses upon natural SARS-CoV-2 infection. This cohort was a subset from a sero-monitoring study previously described (*40*) that was approved by the University of Pennsylvania Institutional Review Board (IRB# 842847). Recent or active SARS-CoV-2 infections were identified based on SARS-CoV-2 RBD antibody levels and/or SARS-COV-2 PCR testing. Longitudinal samples were collected from seropositive participants up to ~200 days post seroconversion to study long-term immune responses. Full cohort and demographic information is provided in **table S1**. Additional healthy donor samples were collected with approval from the University of Pennsylvania Institutional Review Board (IRB# 845061)

Peripheral Blood Sample Processing

Venous blood was collected into sodium heparin and EDTA tubes by standard phlebotomy. Blood tubes were centrifuged at 3000rpm for 15 minutes to separate plasma. Heparin and EDTA plasma were stored at -80°C for downstream antibody analysis. Remaining whole blood was diluted 1:1 with R1 (RPMI + 1% FBS + 2mM L-Glutamine + 100 U Penicillin/Streptomycin) and layered onto SEPMATE tubes (STEMCELL Technologies) containing lymphoprep gradient (STEMCELL Technologies). SEPMATE tubes were centrifuged at 1200g for 10 minutes and the PBMC fraction was collected into new tubes. PBMCs were then washed with R1 and treated with ACK lysis buffer (Thermo Fisher) for 5 minutes. Samples were washed again with R1, filtered with a 70μm filter, and counted using a Countess automated cell counter (Thermo Fisher). Aliquots containing 5-10x10^6 PBMCs were cryopreserved in fresh 90% FBS 10% DMSO.

Detection of SARS-CoV-2 Spike- and RBD-Specific Antibodies

Plasma samples were tested for SARS-CoV-2-specific antibody by enzyme-linked immunosorbent assay (ELISA) as described (*16*, *54*). Plasmids encoding the recombinant full-length Spike protein and the RBD were provided by F. Krammer (Mt. Sinai) and purified by nickel-nitrilotriacetic acid resin (Qiagen). ELISA plates (Immulon 4 HBX, Thermo Fisher Scientific) were coated with PBS or 2 ug/mL recombinant protein and stored overnight at 4C. The next day, plates were washed with PBS containing 0.1% Tween-20 (PBS-T) and blocked for 1 hour with PBS-T supplemented with 3% non-fat milk powder. Samples were heat-inactivated for 1 hour at 56C and diluted in PBS-T supplemented with 1% non-fat milk powder. After washing the plates with PBS-T, 50 uL diluted sample was added to each well. Plates were incubated for 2 hours and washed with PBS-T. Next, 50 uL of 1:5000 diluted goat anti-human IgG-HRP (Jackson ImmunoResearch Laboratories) or 1:1000 diluted goat anti-human IgM-HRP (SouthernBiotech) was added to each well and plates were incubated for 1 hour. Plates were washed with PBS-T before 50 uL SureBlue 3,3',5,5'-tetramethylbenzidine substrate (KPL) was added to each well. After 5 minutes incubation, 25 uL of 250 mM hydrochloric acid was added to each well to stop the reaction. Plates were read with the SpectraMax 190 microplate reader (Molecular Devices) at an optical density (OD) of 450 nm. Monoclonal antibody CR3022 was included on each plate to convert OD values into relative antibody concentrations. Plasmids to express CR3022 were provided by I. Wilson (Scripps).

Detection of SARS-CoV-2 Neutralizing Antibodies

293T cells were seeded for 24 hours at 5 X 10^6 cells per 10 cm dish and were transfected using calcium phosphate with 35 μg of pCG1 SARS-CoV-2 S D614G delta18, pCG1 SARS-CoV-2 S B.1.351 delta18 or pCG1 SARS-CoV-2 S B.1.617.2 delta18 expression plasmid encoding a codon optimized SARS-CoV-2 S gene with an 18-residue truncation in the cytoplasmic tail (kindly provided by Stefan Pohlmann). Mutations in pseudovirus constructs are indicated: D614G (WT) = D614G; B.1.351 = L18F, D80A, D215G, R246I, K417N, E484K, N501Y, D614G, A701V; B.1.617.2 = T19R, G142D, del156-157, R158G, L452R, T478K, D614G, P681R, D950N. 12 hours post transfection, cells were fed with fresh media containing 1mM sodium butyrate to increase expression of the transfected DNA. 24 hours after transfection, the SARS-CoV-2 Spike expressing cells were infected for 2 hours with VSV-G pseudotyped VSVΔG-RFP at an MOI of ~1. Virus containing media was removed and the cells were re-fed with media without serum. Media containing the VSVΔG-RFP SARS-CoV-2 pseudotypes was harvested 28-30 hours after infection, clarified by centrifugation twice at 6000g, then aliquoted and stored at -80 °C until used for antibody neutralization analysis. All sera were heat-inactivated for 30 minutes at 55 ^{0}C prior to use in the neutralization assay. Vero E6 cells stably expressing TMPRSS2 were seeded in 100 μl at 2.5x10^4 cells/well in a 96 well collagen coated plate. The next day, 2-fold serially diluted serum samples were mixed with VSVΔG-RFP SARS-CoV-2 pseudotype virus (100-300 focus forming units/well) and incubated for 1 hour at 37 ^{0}C. 1E9F9, a mouse anti-VSV

Indiana G, was also included in this mixture at a concentration of 600 ng/ml (Absolute Antibody, Ab01402-2.0) to neutralize any potential VSV-G carryover virus. The serum-virus mixture was then used to replace the media on VeroE6 TMPRSS2 cells. 22 hours post-infection, the cells were washed and fixed with 4% paraformaldehyde before visualization on an S6 FluoroSpot Analyzer (CTL, Shaker Heights OH). Individual infected foci were enumerated and the values were compared to control wells without antibody. The focus reduction neutralization titer 50% ($FRNT_{50}$) was measured as the greatest serum dilution at which focus count was reduced by at least 50% relative to control cells that were infected with pseudotype virus in the absence of human serum. $FRNT_{50}$ titers for each sample were measured in at least two technical replicates and were reported for each sample as the geometric mean of the technical replicates.

Detection and Phenotyping of SARS-CoV-2-Specific Memory B Cells

Antigen-specific B cells were detected using biotinylated proteins in combination with different streptavidin (SA)-fluorophore conjugates as described (*16*). All reagents are listed in **table S4**. Biotinylated proteins were multimerized with fluorescently labeled SA for 1 hour at 4C. Full-length Spike protein was mixed with SA-BV421 at a 10:1 mass ratio (200ng Spike with 20ng SA; ~4:1 molar ratio). Spike RBD was mixed with SA-APC at a 2:1 mass ratio (25ng RBD with 12.5ng SA; ~4:1 molar ratio). Biotinylated influenza HA pools were mixed with SA-PE at a 6.25:1 mass ratio (100ng HA pool with 16ng SA; ~6:1 molar ratio). Influenza HA antigens corresponding with the 2019 trivalent vaccine (A/Brisbane/02/2018/H1N1, B/Colorado/06/2017) were chosen as a historical antigen and were biotinylated using an EZ-Link Micro NHS-PEG4 Biotinylation Kit (Thermo Fisher) according to the manufacturer's instructions. Excess biotin was subsequently removed from HA antigens using Zebra Spin Desalting Columns 7K MWCO (Thermo Fisher) and protein was quantified with a Pierce BCA Assay (Thermo Fisher). SA-BV711 was used as a decoy probe without biotinylated protein to gate out cells that non-specifically bind streptavidin. All experimental steps were performed in a 50/50 mixture of PBS + 2% FBS and Brilliant Buffer (BD Bioscience). Antigen probes for Spike, RBD, and HA were prepared individually and mixed together after multimerization with 5uM free D-biotin (Avidity LLC) to minimize potential cross-reactivity between probes. For staining, $5x10^6$ cryopreserved PBMC samples were prepared in a 96-well U-bottom plate. Cells were first stained with Fc block (Biolegend, 1:200) and Ghost 510 Viability Dye for 15 minutes at 4C. Cells were then washed and stained with 50uL antigen probe master mix containing 200ng Spike-BV421, 25ng RBD-APC, 100ng HA-PE, and 20ng SA-BV711 decoy for 1 hour at 4C. Following incubation with antigen probe, cells were washed again and stained with anti-CD3, anti-CD19, anti-CD20, anti-CD27, anti-CD38, anti-CD71, anti-IgD, anti-IgM, anti-IgG, and anti-IgA for 30 minutes at 4C. After surface stain, cells were washed and fixed in 1% PFA overnight at 4C. Antigen-

specific gates for B cell probe assays were set based on healthy donors stained without antigen probes (similar to an FMO control) and were kept the same for all experimental runs.

Detection of Variant RBD, NTD, and S2-Specific Memory B Cells

Variant RBD, NTD, and S2-specific memory B cells were detected using a similar approach as described above. SARS-CoV-2 nucleocapsid was used as a vaccine-irrelevant antigen control. All reagents are listed in **table S4**. Probes were multimerized for 1.5 hours at the following ratios (all ~4:1 molar ratios calculated relative to the streptavidin-only component irrespective of fluorophore): 200ng full-length Spike protein was mixed with 20ng SA-BV421, 30ng N-terminal domain was mixed with 12ng SA-BV786, 25ng wild-type RBD was mixed with 12.5ng SA-BB515, 25ng B.1.1.7 RBD was mixed with 12.5ng SA-BV711, 25ng B.1.351 RBD was mixed with 12.5ng SA-PE, 25ng B.1.617.2 was mixed with 12.5ng SA-APC, 50ng S2 was mixed with 12ng SA-BUV737, 50ng nucleocapsid was mixed with 14ng SA-BV605. 12.5ng SA-BUV615 was used as a decoy probe. All antigen probes were multimerized separately and mixed together with 5uM free D-biotin. Prior to staining, total B cells were enriched from $20x10^6$ cryopreserved PBMC samples by negative selection using an EasySep human B cell isolation kit (STEMCELL, #17954). B cells were then prepared in a 96-well U-bottom plate and stained with Fc block and Ghost 510 Viability Dye as described above. Cells were washed and stained with 50uL antigen probe master mix for 1 hour at 4C. After probe staining, cells were washed again and stained with anti-CD3, anti-CD19, anti-CD27, anti-CD38, anti-IgD, and anti-IgG for 30 minutes at 4C. After surface stain, cells were washed and fixed in 1X Stabilizing Fixative (BD Biosciences) overnight at 4C.

For sorting, pre-enriched B cells were stained with Fc block and Ghost 510 Viability Dye, followed by full-length Spike, WT RBD, and B.1.351 RBD probes as described above. Cells were then stained for surface markers with anti-CD19, anti-CD20, anti-CD27, and anti-CD38, and anti-IgD. After surface stain, cells were washed and resuspended in PBS + 2% FBS for acquisition.

In Vitro Differentiation of Memory B Cells to Antibody Secreting Cells

Memory B cells from bulk PBMC samples were differentiated into antibody secreting cells as described (*39*). Briefly, $1x10^6$ cryopreserved PBMCs were seeded in 1mL of complete RPMI media (RPMI + 10% FBS + 1% Pen/Strep) in 24-well plates. PBMCs were then stimulated with 1000U/mL recombinant human IL-2 and 2.5ug/mL R848 for 10 days. Supernatants were collected at the indicated timepoints. anti-Spike IgG was quantified using a Human SARS-CoV-2 Spike (Trimer) IgG ELISA Kit (Invitrogen) according to the manufacturer's instructions. RBD-ACE2 binding inhibition was measured using a SARS-CoV-2 Neutralizing Ab ELISA Kit (Invitrogen). For anti-Spike IgG experiments, culture supernatants were tested

at 1:100 and 1:1000 dilutions. For RBD inhibition experiments, culture supernatants were tested without dilution and at a 1:2 dilution. Pseudovirus neutralization titers were also measured in culture supernatants starting at a 1:2 dilution as described above.

Detection of SARS-CoV-2-Specific T Cells

SARS-CoV-2-specific T cells were detected using an activation induced marker assay. All reagents are listed in **table S5**. PBMCs were thawed by warming frozen cryovials in a 37°C water bath and resuspending cells in 10mL of RPMI supplemented with 10% FBS, 2mM L-Glutamine, 100 U/mL Penicillin, and 100 ug/mL Streptomycin (R10). Cells were washed once in R10, counted using a Countess automated cell counter (Thermo Fisher), and resuspended in fresh R10 to a density of $5x10^6$ cells/mL. For each condition, duplicate wells containing $1x10^6$ cells in 200uL were plated in 96-well round-bottom plates and rested overnight in a humidified incubator at 37°C, 5% CO2. After 16 hours, CD40 blocking antibody (0.5ug/mL final concentration) was added to cultures for 15 minutes prior to stimulation. Cells were then stimulated for 24 hours with costimulation (anti-human CD28/CD49d, BD Biosciences) and peptide megapools (CD4-S for all CD4+ T cell analyses, CD8-E for all CD8+ T cell analyses) at a final concentration of 1 ug/mL. Peptide megapools were prepared as previously described (*51*, *52*). Matched unstimulated samples for each donor at each timepoint were treated with costimulation alone. 20 hours post-stimulation, antibodies targeting CXCR3, CCR7, CD40L, CD107a, CXCR5, and CCR6 were added to the culture along with monensin (GolgiStop, BD Biosciences) for a 4-hour stain at 37°C. After 4 hours, duplicate wells were pooled and cells were washed in PBS supplemented with 2% FBS (FACS buffer). Cells were stained for 10 minutes at room temperature with Ghost Dye Violet 510 and Fc receptor blocking solution (Human TruStain FcX™ , BioLegend) and washed once in FACS buffer. Surface staining for 30 minutes at room temperature was then performed with antibodies directed against CD4, CD8, CD45RA, CD27, CD3, CD69, CD40L, CD200, OX40, and 41BB in FACS buffer. Cells were washed once in FACS buffer, fixed and permeabilized for 30 minutes at room temperature (eBioscience™ Foxp3 / Transcription Factor Fixation/Permeabilization Concentrate and Diluent), and washed once in 1X Permeabilization Buffer prior to staining for intracellular IFN-g overnight at 4°C. Cells were then washed again and resuspended in 1% paraformaldehyde in PBS prior to data acquisition.

All data from AIM expression assays were background-subtracted using paired unstimulated control samples. For memory T cell and helper T cell subsets, the AIM+ background frequency of non-naïve T cells was subtracted independently for each subset. AIM+ cells were identified from non-naïve T cell populations. AIM+ CD4+ T cells were defined by co-expression of CD200 and CD40L. AIM+ CD8+ T

cells were defined by a Boolean analysis identifying cells expressing at least four of five markers: CD200, CD40L, 41BB, CD107a, and intracellular IFN-g.

Flow Cytometry and Cell Sorting

Samples were acquired on a BD Symphony A5 instrument. Standardized SPHERO rainbow beads (Spherotech) were used to track and adjust photomultiplier tubes over time. UltraComp eBeads (Thermo Fisher) were used for compensation. Up to $5x10^6$ cells were acquired per sample. Data were analyzed using FlowJo v10 (BD Bioscience). For Boolean analysis of variant cross-binding, data were imported into SPICE 6 (NIH Vaccine Research Center (*55*)). Cell sorting was performed on a BD FACS Aria II instrument in low pressure mode using a 70um nozzle. Cells were sorted into DNA LoBind Eppendorf tubes containing cell lysis buffer (Qiagen).

B Cell Receptor Sequencing

Library Preparation

DNA was extracted from sorted cells using a Gentra Puregene Cell kit (Qiagen, catalog no. 158767). Immunoglobulin heavy-chain family–specific PCRs were performed on genomic DNA samples using primers in FR1 and JH as described previously (*47*, *56*). Two biological replicates were run on all samples. Sequencing was performed in the Human Immunology Core Facility at the University of Pennsylvania using an Illumina 2× 300-bp paired-end kit (Illumina MiSeq Reagent Kit v3, 600-cycle, Illumina MS-102-3003).

IGH Sequence Analysis

Reads from an Illumina MiSeq were filtered, annotated, and grouped into clones as described previously (*16*, *57*). Briefly, pRESTO v0.6.0 (*58*) was used to align paired end reads, remove short and low-quality reads, and mask low-quality bases with *N*s to avoid skewing SHM and lineage analyses. Sequences which passed this process were aligned and annotated with IgBLAST v1.17.0 (*59*). The annotated sequences were then imported into ImmuneDB v0.29.10 (*60*, *61*) for clonal inference, lineage construction, and downstream processing. For clonal inference, sequences with the same IGHV gene, IGHJ gene, and CDR3 length from each donor were hierarchically clustered. Sequences with 85% or higher similarity in their CDR3 amino-acid sequence were subsequently grouped into clones. Clones with productive rearrangements and >= 2 copies were filtered for downstream analysis.

Lineage Construction & Visualization

For each clone, a lineage was constructed with ImmuneDB as described in (*61*). ete3 (*62*) was used to visualize the lineages where each node represents a unique sequence, the size of a node represents its relative copy number fraction in the clone, and the integer next to each node represents the number of mutations from the preceding vertical node.

Overlapping Clone SHM Analysis

Clones were filtered based on size using a copy number filter such that clones which had a copy number less than 50% of the mean copy number frequency (50% mcf) within the subject were excluded. From this population, only clones that appeared in both WT RBD and cross-binder (RBD++) samples were included. The SHM of each clone was averaged across each unique sequence, weighted by the copies of each sequence, and visualized as categorical variables (pie chart) and as frequencies (boxplots).

Data Availability

Raw sequencing data for all donors and subsets is available on SRA under BioProject PRJNA752617. Processed AIRR-seq data will be made available on the AIRR Data Commons via the iReceptor portal (*63*).

<u>Estimating Decay Rates</u>

To understand and compare the rate of loss of immune responses after vaccination, we tested different statistical models of decay against the data. We first tested if there was significant decay (i.e. was the decay rate significantly different from zero). We then tested if there was evidence for a slowing of decay with time (using a two-phase model). This is a heuristic approach to understanding decay and does not imply a mechanism or that the underlying immune dynamics may be more complex. The decay rate post-second dose of vaccine was estimated using a censored mixed effect regression framework. Briefly, the dependency of variables of interest on days post vaccine can be modelled by using either one constant decay slope or a decay slope that changes with time (assume a two- phase decay with a fixed break point at T_0). The model of the immune response y for participant i at time t_{ij} can be written as below:

$$y_{ij} = \beta_0 + b_{0i} + \beta_1 t_{ij} + b_{1i} t_{ij}$$ – for a model with a single slope; and

$$y_{ij} = \beta_0 + b_{0i} + \beta_1 t_{ij} + b_{1i} t_{ij} + \beta_2 s_{ij} + b_{2i} s_{ij}$$ – for a model with two different slopes, in which:

$$s_{ij} = \begin{cases} 0, & t_{ij} < T_0 \\ t_{ij} - T_0, & t_{ij} \geq T_0. \end{cases}$$

The parameter β_0 is a constant (global intercept), and b_{0i} is a patient-specific adjustment (random effect) to the global intercept. The slope parameter β_1 is a fixed effect to capture the average decay rate for all

individuals before T_0; and b_{1i} is a patient-specific random effect of the decay rate. To fit the model with a two-phase decay slope (with break point at time T_0), an extra parameter β_2 (with a subject-specific random effect b_{2i}) was added to represent the difference between the two slopes. Throughout the manuscript, we chose the median of the timepoints post-second dose of vaccine as the break point in decay rate (i.e., T_0=day 89).

To account for values less than the detection threshold in the assay, a censored mixed-effect regression method was used to estimate the parameters in the model. Values less than 10 were censored for the neutralization data. For T cell measurements, this detection threshold varies (see supplemental information – determining the limit of detection for details on how this variable limit of detection was captured). The linear models above were fitted with censoring of values below the limit of detection using lmec library in R (*64*) (with the maximum likelihood algorithm option to fit for the fixed effects). We used a likelihood ratio test to determine if the response variables where better fit with either the single or two-phase decay models (by testing whether $\beta_2 = 0$), and to test whether the decay rates were different between SARS-CoV-2 naïve and recovered subjects (this test compares the likelihood value of the nested models and the difference in the number of parameters). These analyses were carried out in R version 4.0.4.

Determining the Limit of Detection for Estimating Decay Rates

For each individual and at each time point (i.e. each sample) the limit of detection in assays of T cell stimulation varied. This is because the background level is determined by running paired assessment of cells from a given sample in (SARS-CoV-2 peptide) stimulated and unstimulated cultures. The quantify of interest (of which we wish to measure the decay rate) is the difference in the fraction of T cells activated in the stimulated and unstimulated cultures. The variable limit of detection (LOD) for each sample must be considered when determining the decay rate for T cell responses. To determine if the fraction of activated cells in a stimulated sample was significantly higher than the fraction of activated cells in the corresponding unstimulated sample (i.e. if the sample was above the limit of detection) we used a one-sided two proportion Z test. Formally, we let the proportion of unstimulated and stimulated responses (over total non-naïve cells) be denoted by $U_{i,j}$ and $S_{i,j}$ for patient i at time j, respectively. It follows that we are interested in estimating the decay rate of the quantity $\Delta_{i,j} = S_{i,j} - U_{i,j}$. A one-sided two proportion Z test was used to determine if $S_{i,j} > U_{i,j}$. Briefly, for each patient i at time j, the following quantity was calculated:

$$Z_{i,j} = \frac{\Delta_{i,j}}{\sqrt{p(1-p)\left(\frac{1}{n_{s_{i,j}}} + \frac{1}{n_{u_{i,j}}}\right)}}$$

With:

$\Delta_{i,j} = S_{i,j} - U_{i,j}$,

$p = \frac{S_{i,j} \times n_{s_{i,j}} + U_{i,j} \times n_{u_{i,j}}}{n_{s_{i,j}} + n_{u_{i,j}}}$,

$n_{s_{i,j}}$ = total non-naïve cells in stimulated group for subject i at time j,

$n_{u_{i,j}}$ = total non-naïve cells in unstimulated group for subject i at time j.

For each subject, we calculated the minimum difference needed to achieve significance by solving the above equation for $\Delta_{i,j}$ (assuming p is constant) at the $Z_{critical}$ level (ie, with α=0.05, $Z_{critical}$=1.645 for a one-sided test). This minimum difference can be written as:

$$\Delta_{MIN\,i,j} = 1.645 \times \sqrt{p(1-p)\left(\frac{1}{n_{s_{i,j}}} + \frac{1}{n_{u_{i,j}}}\right)}$$

We censored subject i if the difference is not statistically significant (ie, $Z_{i,j}$<1.645, with α=0.05). The detection limit for subject i was calculated by taking the maximum value of $\Delta_{MIN\,i,j}$ across all timepoint for that subject. The values $\Delta_{i,j}$ were normalized by the maximum $\Delta_{MIN\,i,j}$ for each subject, hence the limit of detection was set to zero, and the lmec regression models applied to the normalized data in order to determine the decay rates of T cell responses.

High Dimensional Analysis and Statistics

All data were analyzed using custom scripts in R and visualized using RStudio. Pairwise correlations between variables were calculated and visualized as a correlogram using corrplot with FDR correction as described previously (*65*). For heatmaps, data were visualized with pheatmap. For construction of UMAPs, 12 antigen-specific immune features were selected: anti-Spike IgG, anti-RBD IgG, D614G FRNT50, B.1.351 FRNT50, Spike+ memory B, RBD+ memory B, % IgG+ of Spike+ memory B, % IgG+ of RBD+ memory B, AIM+ CD4 T, AIM+ CD4 Tfh, AIM+ CD4 Th1, and AIM+ CD8 T. Antibody and cell frequency data were log10 transformed and scaled by column (z-score normalization) prior to generating UMAP coordinates. Statistical tests are indicated in the corresponding figure legends. All tests were performed two-sided with a nominal significance threshold of $p < 0.05$. Benjamini-Hochberg (BH) correction was performed in all cases of multiple comparisons. Unpaired tests were used for comparisons between timepoints unless otherwise indicated as some participants were missing samples from individual timepoints. * indicates $p < 0.05$, ** indicates $p < 0.01$, *** indicates $p < 0.001$, **** indicates $p < 0.0001$. Source code and data files are available upon request from the authors.